УДК 004.056, 004.77

ОРГАНИЗАЦИЯ МОНИТОРИНГА В ТЕЛЕКОММУНИКАЦИОННЫХ СЕТЯХ С ЦЕЛЬЮ ОБНАРУЖЕНИЯ ИНФОРМАЦИОННЫХ УГРОЗ БЕЗОПАСНОСТИ ПЕРЕДАЧИ ДАННЫХ

А.В. Гирик

Рассмотрены современные подходы к анализу трафика в телекоммуникационных сетях, их достоинства и недостатки. Предлагается модель обнаружения информационных угроз безопасности передачи данных в сетях путем построения поисковых прогнозов показателей безопасности и сравнения их с нормальным профилем сети.

Ключевые слова: мониторинг вычислительных сетей, анализ и прогнозирование трафика, обнаружение информационных угроз безопасности передачи данных.

Введение

Обеспечение информационной безопасности – одна из наиболее актуальных и сложных проблем в современных сетях передачи данных [1]. Очевидно, что она должна решаться комплексно, так как:

1. сеть представляет собой совокупность взаимодействующих программных и аппаратных средств, построенных в соответствии с множеством различных стандартов различными производителями, т.е. вычислительную среду с высокой степенью гетерогенности. Специфика каждого сетевого элемента должна учитываться при проектировании системы управления безопасностью сети;

2. сетевое взаимодействие осуществляется по протоколам, принадлежащим определенному сетевому стеку, например, TCP/IP или OSI. Необходимо заботиться о том, чтобы на каждом уровне была обеспечена надежная защита передаваемых данных;
3. наличие разнородных механизмов защиты в сети, их взаимное влияние и влияние на производительность сети должны тщательно исследоваться. Работа различных подсистем безопасности должна быть согласована, что на практике, однако, достигается не всегда [2];
4. при проектировании системы защиты необходимо ориентироваться на статистику угроз и учитывать возможность появления потенциальных (неизвестных) угроз.

Безопасность сети определяется защищенностью всех входящих в нее компьютеров и сетевого оборудования, и злоумышленнику в некоторых случаях может быть достаточно нарушить работу только одного компонента, чтобы скомпрометировать всю сеть [3]. Обеспечение безопасности должно решаться как часть задачи управления сетью. Как правило, эта задача возлагается на специализированный комплекс программных средств, который называется системой управления сетью (Network Management System, NMS) [4].

Система управления сетью берет на себя функцию сбора и анализа сведений о поведении трафика в сети. Это делается, в том числе, и для того, чтобы иметь возможность прогнозировать значения трафика узлов сети. Функции построения прогнозов реализуются в дорогих системах управления сетью в качестве дополнительных модулей. В контексте сетевой безопасности прогнозы используются в основном для обнаружения сетевых аномалий, например, нештатных ситуаций в работе сети, сетевых перегрузок, атак (типа «отказ в обслуживании» и др.) [5].

Мониторинг вычислительных сетей

Рассмотрим кратко современные технологии и подходы к сетевому мониторингу. К числу получивших наибольшее распространение, в том числе использованных во множестве научных исследований, можно отнести:

- мониторинг на основе Simple Network Management Protocol (SNMP);
- мониторинг потоков данных (NetFlow, JFlow, NetStream, sFlow);
- анализ сетевых пакетов;
- трассировка событий сетевого стека (например, с помощью DTrace);
- сквозной мониторинг (на основе ICMP, UDP, TCP);
- инструментирование приложений;
- анализ журналов прикладных и системных программ.

SNMP является стандартным средством сбора информации и управления сетевыми устройствами в сетях TCP/IP [6]. Он работает по схеме менеджер–агент и позволяет управляющей программе (менеджеру) взаимодействовать с агентом (приложением, которое, как правило, входит в состав системного ПО элемента сети, например, коммутатора или маршрутизатора) с целью получить данные о работе устройства или изменить конфигурацию устройства. Все параметры работы устройства содержатся в базе данных, логически организованной в виде дерева. Среди них можно выделить параметры, определенные в одном из стандартов базы (например, MIB-II или RMON2), и параметры, уникальные для каждого типа устройства. Для работы с SNMP существует большое количество коммерческих и бесплатных программ, в частности, пакеты net-snmp и OpenSNMP.

Мониторинг на основе потоков данных приобрел популярность после того, как Cisco Systems реализовала в своих маршрутизаторах протокол NetFlow, позволяющий управляющему приложению (коллектору) получать сведения о потоках IP-пакетов, обработанных устройством. Поток определяется как однонаправленная последователь-

ность пакетов, у которых совпадают ip-адреса и номера портов источника и получателя и тип протокола [7]. Записи протокола NetFlow сохраняют большое количество информации о пакетах, в том числе значение поля ToS, информацию, специфичную для транспортных протоколов, номер автономной системы и многое другое. Поскольку обработка всех пакетов может приводить к большой нагрузке на сетевое устройство, часто используется сэмплирование (sampled NetFlow), когда сохраняется информация лишь о каждом *n*-ном пакете. В этом случае сохраненные сведения являются лишь оценками реальных потоков. Популярность протокола NetFlow привела к тому, что комитет IETF создал рабочую группу для разработки универсального протокола экспорта информации о потоках пакетов, обрабатываемых сетевыми устройствами (Internet Protocol Flow Information eXport, IPFIX) [8]. Маршрутизаторы Juniper Networks используют похожий протокол JFlow, а устройства Huawei Technologies – NetStream. В сетевых устройствах некоторых производителей, например, D-Link, реализован протокол sFlow, обладающий более широкими возможностями по сравнению с NetFlow [9].

Анализ сетевых пакетов осуществляется с помощью перехватчика пакетов (сниффера). Этот способ позволяет получить любую информацию о проходящих через сетевой стек хоста пакетах, применить сложные алгоритмы фильтрации и поиска нужной информации. Для мониторинга сети в целом он применяется ограниченно в силу следующих причин:

– создает большую нагрузку на систему, резко снижая ее производительность;
– не применим для закрытых устройств типа коммутаторов и маршрутизаторов;
– в сети на основе коммутаторов перевод интерфейса в режим беспорядочного захвата пакетов (promiscuous mode) не позволяет захватывать пакеты, предназначенные другим станциям в сети (хотя эта трудность может быть преодолена с помощью зеркалирования трафика со всех или некоторых портов коммутатора на выделенный порт).

Наиболее популярные снифферы для UNIX-систем – tcpdump и snoop. Многие снифферы используют в качестве основного средства доступа к стеку библиотеку libpcap, портированную на множество платформ, в том числе и MS Windows. Примером приложения, использующего библиотеку libpcap, является сниффер WireShark (ранее Ethereal), названный eWEEK Labs «одним из самых важных open source приложений».

Трассировка событий сетевого стека представляет другой подход к сбору информации о работе элементов сети. Приложения-трассировщики осуществляют мониторинг производимых в системе операций и заносят в журнал временные метки событий, связанных с этими операциями, например, событий установления и разрыва TCP-соединения. Трассировка создает потенциально меньшую нагрузку на систему, чем перехват пакетов, однако в остальном обладает теми же недостатками, что и анализ сетевых пакетов. DTrace является мощным программным комплексом для исследования поведения операционных систем Sun Microsystems с помощью трассировки; в системах, основанных на ядре Linux, может применяться SystemTap, в системах Microsoft – Event Tracing for Windows.

Протокол Internet Control Message Protocol также может использоваться для мониторинга сети, в частности, для получения сквозных показателей функционирования сети. Изменение времени двойного оборота пакетов может свидетельствовать об изменениях в сетевой нагрузке и нагрузке конечного узла. Несмотря на свою популярность в качестве средства тестирования связности, метод на основе ping'а не получил широкого распространения для анализа производительности и построения прогнозов [10]. Для получения оценки времени двойного оборота пакетов могут использоваться также протоколы UDP (например, ответы так называемых well-known services типа ECHO или TIME), TCP, DHCP и вообще любые протоколы, предусматривающие ответ на некоторое сообщение [11].

В случае, если нужна более подробная информация о составляющих времени ответа, необходимо подняться на прикладной уровень и использовать инструментирование приложений [12]. Задача может решаться по-разному в зависимости от того, есть ли доступ к исходным кодам и возможность пересборки приложения или нет. В первом случае предпочтительным способом инструментирования является инструментирование на основе стандартов (например, ARM или AIC), во втором – инструментирование путем внедрения кода или использования возможностей операционной системы (например, отладка) и встроенных в приложение возможностей получения информации о происходящих в нем событиях.

Анализ журналов, или логов, программ (например, анализ логов веб-сервера) может быть полезным источником информации о сетевых событиях, однако в качестве полноценного источника данных о работе сети рассматриваться не может.

Существует ряд способов получения дополнительной информации о событиях в элементах сети, например, использование протокола syslog. Как правило, они применяются для более точной идентификации сетевых аномалий при совокупном анализе с данными по потокам и SNMP-статистикой.

Таким образом, чтобы получить наиболее точную информацию о работе сети, желательно использовать комбинированный подход на основе анализа SNMP-статистики и данных NetFlow [5, 13]. Такая модель находит применение во многих системах обнаружения вторжений в сеть (Network Intrusion Detection System). В общем случае угрозу нужно сначала обнаружить, а затем идентифицировать. Как правило, для идентификации угрозы необходимо больше информации, чем для ее обнаружения, но более информативный мониторинг создает б*о*льшую нагрузку на сеть и приводит к потерям производительности. Поэтому целесообразно задействовать механизмы подробного анализа после обнаружения угрозы.

Анализ данных мониторинга

Рассмотрим подход к работе с агрегированными источниками данных, использованный при анализе работы сети одного из провайдеров Санкт-Петербурга. В первую очередь, необходимо идентифицировать источники данных. В сетях провайдеров в качестве источников данных выступают управляемое сетевое оборудование (коммутаторы, маршрутизаторы) и серверы, на которых выполняются ключевые сервисы (биллинговая система, система сбора статистики, файловые, почтовые сервисы и т.д.). В корпоративных сетях в качестве источника данных следует рассматривать также рабочие станции, на которых можно выполнить инструментирование приложений и осуществлять мониторинг действий пользователя (т.е. установить клиентскую часть системы обнаружения вторжений).

Следующим шагом является консолидация данных, поступающих от различных источников, преобразование данных в формат, пригодный для передачи ядру системы обнаружения вторжений, и их запись в хранилище данных. В тех случаях, когда в сети уже настроен и функционирует сервис, выполняющий мониторинговые функции (например, Cacti или Nagius), или полноценная система управления сетью, целесообразно воспользоваться этим обстоятельством и извлекать данные для анализа из баз данных этих систем.

Наиболее сложной частью является собственно анализ. Для статистической системы обнаружения вторжений можно предложить несколько моделей анализа данных, наиболее перспективной из которых представляется анализ временных рядов на основе авторегресии [14]. Построение прогнозов на основе анализа временных рядов получило значительное распространение в эконометрике и впоследствии стало использоваться и

для моделирования поведения процессов в сетях передачи данных. В общем случае временной ряд может быть представлен мультипликативной моделью вида

$$X(t)=T(t)\cdot C(t)\cdot S(t)\cdot\varepsilon(t), \quad (1)$$

где $T(t)$ – основная закономерность развития процесса во времени, или тренд, $C(t)$ – циклическая составляющая, $S(t)$ – сезонная составляющая, $\varepsilon(t)$ – случайные колебания. Экспериментальные данные показывают, что в большинстве случаев трафик в сети обнаруживает периодические колебания. Анализ автокоррелограммы позволил выявить сезонные эффекты с периодом 1 час, 24 часа (сутки) и 168 часов (неделя).

Расчет прогноза по методу экспоненциального сглаживания и методу Хольта–Уинтерса позволяет судить о большей точности и гибкости последнего, однако подбор коэффициентов адаптации для каждого ряда представляет собой отдельную сложную задачу. Значение экспоненциально сглаженного ряда рассчитывается как

$$\bar{x}_\lambda(t)=\frac{1-\lambda}{1-\lambda^t}\sum_{m=0}^{t-1}\lambda^m x(t-m), \quad (2)$$

где λ – параметр адаптации ($0<\lambda<1$). Метод Хольта–Уинтерса позволяет учесть сезонные эффекты и ослабить ограничения, присущие методу экспоненциального сглаживания, которые связаны с его однопараметричностью. Прогнозное значение рассчитывается как

$$\bar{x}(t,l)=[a(t)+l\cdot b(t)]w(t+l-T), \quad (3)$$

где l – шаг прогноза, w – коэффициент сезонности, T – число временных тактов, содержащихся в сезонном цикле. Сезонность здесь, как и в (2), представлена мультипликативно. Обновление коэффициентов прогнозирования выполнялось по формулам

$$a(t+1)=\lambda_1\frac{x(t+1)}{w(t+1-T)}+(1-\lambda_1)[a(t)+b(t)], \quad (4)$$

$$b(t+1)=\lambda_2[a(t)+b(t)]+(1-\lambda_2)b(t), \quad (5)$$

$$w(t+1)=\lambda_3\frac{x(t+1)}{a(t+1)}+(1-\lambda_3)w(t+1-T), \quad (6)$$

где λ_1, λ_2 и λ_3 – адаптирующие параметры. На рисунке представлены графики средней ошибки прогноза $\bar{e}(l)$, которая на исследованных наборах данных оказалась меньше для метода Хольта–Уинтерса.

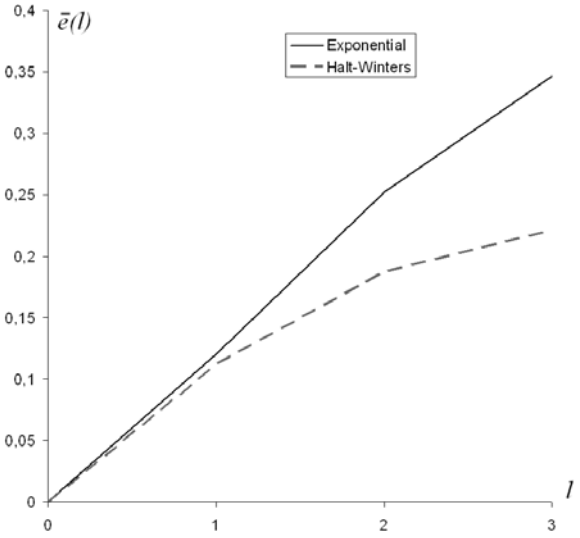

Рисунок. Средняя ошибка прогноза с помощью методов экспоненциального сглаживания и Хольта-Уинтерса на агрегированных данных по трафику в сети

Многошаговое прогнозирование на основе метода экспоненциального сглаживания менее эффективно, так как использует прогнозное значение, рассчитанное на предыдущем шаге, для расчета следующего прогнозного значения, в результате чего быстрее накапливается ошибка. В [13] авторы рекомендуют перейти к аддитивной модели ряда. Это можно сделать, применив, например, аддитивную модель сезонности Тейла–Вейджа.

Нахождение наиболее точного прогноза может решаться как задача оптимизации и заслуживает рассмотрения в отдельной публикации. После того, как прогноз сформирован, его нужно сравнить с нормальным профилем для исследуемого показателя. Нормальный профиль показателя формируется на данных, аккумулированных ко времени построения прогноза, для реальных систем желательно использовать профиль, построенный на данных за несколько месяцев. В эксперименте использовались данные по трафику в сегменте сети, собранные с коммутаторов второго уровня, за период с марта по сентябрь 2008 г.

По результатам сравнения делается вывод о наличии критического отклонения. Если критическое отклонение зафиксировано для нескольких источников данных, становится возможным предпринять действия, которые позволят предотвратить атаку или вовремя информировать администратора сети о возникновении нештатной ситуации.

Заключение

Предложен подход к разработке системы обнаружения вторжений на основе построения поисковых прогнозов показателей безопасности и сравнения их с нормальным профилем сети. Построение многошаговых прогнозов на основе модели Хольта–Уинтерса позволяет учесть сезонность и получить низкую по сравнению с экспоненциальным методом сглаживания ошибку для двух и более шагов. Если отклонение от ожидаемых значений лежит вне допустимых границ, то нестандартное поведение показателя рассматривается как признак возможной атаки или возникновения нештатной ситуации.

Литература

1. Жигулин Г.П. Прогнозирование устойчивости субъектов информационного взаимодействия. – СПб: СПбГУ ИТМО, 2006. – 191 с.
2. V. Paxson Bro: A system for detecting network intruders in real-time // Computer Networks. – 1999. – V. 31. – P. 2435–2463,
3. Зегжда Д.П., Ивашко А.М. Основы информационной безопасности информационных систем. – М.: Горячая линия – Телеком, 2000. – 449 с.
4. Galis A. Multi-Domain Communication Management. – CRC Press, 2000 – 420 pp.
5. Barford P., Plonka D. Characteristics of Network Traffic Flow Anomalies // Proceedings of ACM SIGCOMM Internet Measurement Workshop, San Francisco, 2001.
6. Harrington D., Presuhn R., Wijnen B. RFC 3411: An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks, 2002. – Режим доступа: http://tools.ietf.org/html/rfc3411, свободный.
7. CAIDA : analysis : workload : flowtypes [Электронный ресурс]. – Режим доступа: http://www.caida.org/analysis/workload/flowtypes/, свободный.
8. Claise B. RFC 5101: Specification of the IP Flow Information Export (IPFIX) Protocol, IETF Proposed Standard, 2008 [Электронный ресурс]. – Режим доступа: http://tools.ietf.org/html/rfc5101, свободный.
9. Phaal P., Lavine M. sFlow version 5, 2004 [Электронный ресурс]. – Режим доступа: http://www.sflow.org/sflow_version_5.txt, свободный.
10. Internet End-to-end Performance Monitoring [Электронный ресурс]. – Режим доступа: http://www-iepm.slac.stanford.edu/, свободный.

11. Well-known services list [Электронный ресурс]. – Режим доступа: http://www.watchguard.com/help/lss/46/reference/ports4.htm, свободный.
12. Гирик А.В. Инструментирование клиент-серверных приложений // Научно-технический вестник СПбГУ ИТМО. – 2005. – № 19. – С. 150–154.
13. Papadopouli M., Shen H., Raftopoulos E., Ploumidis M., Hernandez-Campos F. Short-term traffic forecasting in a campus-wide wireless network. – 16th Annual IEEE International Symposium on Personal Indoor and Mobile Radio Communications, Berlin, 2005.
14. Гирик А.В. Обнаружение информационных угроз безопасности передачи данных в телекоммуникационных сетях // Труды XV Всероссийской научно-методической конференции «Телематика'2008». – 2008. – С. 178–179.

Гирик Алексей Валерьевич — Санкт-Петербургский государственный университет информационных технологий, механики и оптики, преподаватель, alexei.guirik@googlemail.com